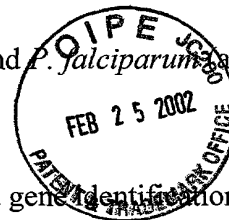


number of exons, clustered at the right side of the presented sequence), and *P. falciparum* (a simple gene, in blue).



[0106] Following such analyses, it appears that the physics-based gene identification method is not necessarily very accurate in the detection of the various exons in a gene, such as the one above for *T. arabidopsis*, when the stability regions corresponding to the different exons are not sharply distinguished from each other. On the other hand, the physics-based method is very powerful in the straightforward identification of genes for which the exons are sharply discriminated from each other, such as for the examples above in *C. elegans* and *D. melanogaster*.

[0107] The above result can then be used for the practical discovery of potential new genes in a genome such as *Homo sapiens*. As an illustration, we consider below a genomic sequence from *H. sapiens* (Accession AP001754). In the original annotation, genes are reported only for the first half of the genomic sequence (length 340,000 bp), and these genes are represented in red and blue. See Fig. 16.

[0108] With the physics-based gene identification scheme, potential new genes are discovered in the is second half, as represented partially (exons in red, in the region 270 to 300 kbp, and further zooming as shown in Fig. 17). Part of the above analysis is presented in Figure 21 in more detail. In the original genomic sequence the coding regions are in green, regions as discovered by the physics-based are highlighted in blue text (the non-coding regions are in green, and the splice sites in magenta).

2022-10-20 09:50:50

This page intentionally left blank

Eukaryotic genomes, particular case of *Anopheles gambiae*:

[0109] For the gene identification procedure as described above, it appears that the genomes of insects represent particularly favorable cases, with possibly the identification of all genes (and almost all exons). This situation is illustrated with *D. melanogaster* and *A. gambiae*, for example.

[0110] In the case of *A. gambiae*, notably, with the conditions adopted throughout this document, it appears that very few temperatures (74°C, 75°C, and 76°C with the conditions chosen, and described above) could be enough as a matter of fact for the proper detection of genes and exons. This feature is illustrated below with the gene AgProPO (Accession Number: AF031626).